Data Science Consulting

Statistics 415 / 515 • Spring 2018 • Genevera Allen

Time & Location:

Clinic: Mondays 1:00pm - 4:00pm	Location: RZR 113
Lecture: Thursdays 4:00 - 5:00pm	Location: AEL A121

Instructor:

Genevera Allen
Office: Duncan Hall 2080
Email: gallen@rice.edu
Office Hours: By appointment

Course Description:

Students in this course will advise clients in a data science consulting clinic, learn best practices in consulting, and gain exposure to a variety of real data science problems.

Course Overview:

The main component of this course will be a data science consulting clinic where students will advise clients with data science problems from across the Rice community. During the consulting clinic, the instructor will quickly assess the needs of each client and assign them to a small group of students with relevant expertise for help. The instructor will then supervise the student groups and provide input throughout the clinic. After the clinic, student groups will be required to write up a standardized report for each client they interacted with. During an additional lecture hour, the class will meet to review problems brought into the clinic each week and learn best practices in data science consulting.

Course Objectives & Outcomes:

Students will learn best practices in data science consulting and be exposed to a variety of real data science problems. Students completing this course should be able to (i) listen and ask questions of clients to ascertain the data science challenges, (ii) formulate challenges into coherent data science problems, (iii) provide guidance or simple solutions on the identified problems, and (iv) communicate with clients to help them understand the problems, solutions, and how to interpret results.

Prerequisites & Background:

Students should have taken relevant data science courses and have familiarity in the following areas and expertise in at least one area: Data Acquisition, Data Wrangling, Data Visualization, Databases and Data Management, Statistical Analysis, Statistical Modeling, and Machine Learning. Students should also have excellent programming skills (in a language such as R or Python) and some prior experience working with real data science problems.

Application Process:

Applications should be sent to Genevera Allen (gallen@rice.edu) by 5pm on **Monday, November 6th**. Applications should include:

1. A one page personal statement. In this statement, please discuss why you want to enroll in data science consulting and what is your relevant background and experience. Please discuss any relevant coursework, programming skills, and your experience with real-world data science projects. Also, specifically discuss your skills and experience in each of the following areas: Data Acquisition, Data Wrangling, Data Visualization, Databases and Data Management, Statistical Analysis, Statistical Modeling, and Machine Learning.

2. A current CV or resume.

3. An example of a real-world data science project completed as part of a course, internship, or research project. Feel free to submit a written report, software package, or other representative work of this project.

Applications will be evaluated by a group of three faculty members in the Statistics and Computer Science department. Students will be selected for the course based upon their relevant experience as well as to ensure that the selected students have a diverse set of skills that represent all aspects of data science. Students will be notified

about their selection for the course by Friday, November 10th, giving them time to enroll when registration opens on November 13th. Fifteen students will be selected for the course - five graduate students enrolled in 515 and ten undergraduates enrolled in 415.

Course Organization & Schedule:
The first two weeks of the course will include a crash course on data science consulting as well as outreach to advertise the clinic to potential clients. Specifically, the first two weeks of clinic time will include lectures on basic consulting skills and the process of data science consulting, visits by data science consultants working in industry, and visits by potential clients for model consulting interactions and mock consulting sessions. Additionally, students will be asked to advertise the data science consulting clinic at Rice by posting fliers and sending advertising emails to departments across campus.

The data science consulting clinic will open the third week of classes and continue through the end of the semester. If too few clients come to the clinic, we will open up and advertise the consulting clinic to TMC and UH researchers. Additionally, the instructor will review data science consulting case studies (amassed from several years of TMC consulting office hours) in any down time in the clinic.

During the lecture hour, students will briefly present case studies from the clinic. These will be discussed as a class to decide on the best recommendations for the clients and also so that other students can learn about new problems.

Coursework:
The major requirement of this course is to attend and participate as a consultant in the clinic. Each team of students working with a client will also be required to write a standardized report and present a brief case study in the lecture hour. The lead consultant will take primary responsibility for the report and presentation. (Lead consultants will either be assigned by the instructor or chosen by the student groups.)

Grading:
Attendance: 25%
Consulting evaluations: 15%
Written consulting reports: 30%
Case study presentations: 25%
Outreach & Client Recruitment: 5%
A separate document will cover the specific format and requirements of the written reports as well as detailed grading rubrics. Consulting evaluations will include a combination of peer evaluations by other students on the consulting team as well as instructor evaluations; a separate document will include a rubric for evaluations.

Stat 415 vs. 515:
Graduate students should enroll in Stat 515 and undergrads in Stat 415. Graduate students will serve as lead consultants who help manage interactions with the clients and are responsible for coordinating the consultant team's written report and oral presentation.

Book-keeping:
Records will be kept on all clients that come to the clinic as part of the standardized reports student teams write. Additionally, every client will be emailed after the clinic and asked to complete a short survey about their experiences. Survey results will be used to continually improve the data science consulting clinic throughout the semester.

Students with a documented disability requiring accommodations should speak with the instructor during the first two weeks of class.

This syllabus is subject to change with reasonable advance notice by the instructor.